

# National Library of Korea's Digitization Service



# Contents

## I What We Do 03

Korean Memory establishment

National collection digitization and service

## II How We Do the Work 08

National collection digitization process

Key success factors of national collection digitization

## III Future tasks 17

Opening, sharing and expanding digitized materials

Case 1) East Asia Digital Library(EADL)

Case 2) Korea Newspaper Archive

Case 3) Japanese Government General Gazette Utilization System

Case 4) E-picture Card Service



## 「Korean Memory」, the library of memories

Culture Vision 2030  
(New Culture Policy Preparation Group New  
Culture Policy Planning Team of the  
MCST, 2018. 5. 16.)

- **Agenda7. Reinforce culture resource convergence capability**
- **Agenda4. Strengthen culture resource record/preservation system**

(expand archive) increase support for national/public/private culture institutions' archiving platform to improve culture resource recording, preservation and utilization system

As part of the Korean Memory, the library of memories project\*

- NLK mid- to- long- term development plan implementation task 1- 2- 1
- 「NLK future strategy TF report」 digital/technology area implementation task 3 - 1
  - Build national digital collection by collecting/preserving knowledge resources that hold collective memory of culture art institutions and individuals
  - Support digital knowledge culture heritage archive platform

Need to develop and run archive system for atypical digital archive material\* collection/preservation/organization/ service



\* Photos, manuscripts, letters, oral sound sources before publication(texts, images, sound sources, video clips)



# Korean Memory establishment

## 「Korean Memory」, the library of memories

Continuously implement high quality/mass digitization of national collection

- Implement systematic high quality/mass digitization of the NLK's collection and rare materials of culture and art institutions to expand content.
- Continuously implement quality management.

Digital curation plan

- Organize collection by year/subject/person/incident to successfully carry out Korean Memory
- Strengthen curation to allow tridimensional access to contents

Define digital content creation guideline and standard data

- Establish and manage resource creation guideline to define standard data and digitize contents by form and media

Develop content management system and service platform

- Build content management system (link with participating institutions' data etc.)
- Create a model by predicting user demand
- Based on the model, develop a service platform with a user-centered interface

Establish and run the National Association for Knowledge Culture Heritage

- Increased participation of national and institutions under MCST including the National Museum of Korea, National Folklore Museum, National Museum of Modern and Contemporary Art and national/public institutions including the Cultural Heritage Administration, National Institute of Korean History, Independence Hall of Korea, Korea Democracy Foundation



# National Collection Digitization and Service

## The NLK's task

### Representative library of Korea (Library Act Article 18)

#### Library Act Article 19

Comprehensive/collection/provision/preservation of library materials

National bibliography compilation/standardization/digitization



Support for/cooperation with libraries in Korea

Exchange and cooperation with foreign libraries

Reading promotion activity support and cooperation

# National Collection Digitization and Service

Digitization and full-text service

## “ Digitization and use of the library’s collection (Library Act, Copyright Act) ”

Library Act Article 20, Article 44

Collect/provide/preserve  
materials through legal  
deposit

Reduce knowledge  
information gap



Copyright Act Article 31

Replication in digital form for  
library collection  
preservation

Transmission of digitized  
materials within the library

Transmission of digitized  
materials between libraries

# National Collection Digitization and Service

## Digitization project

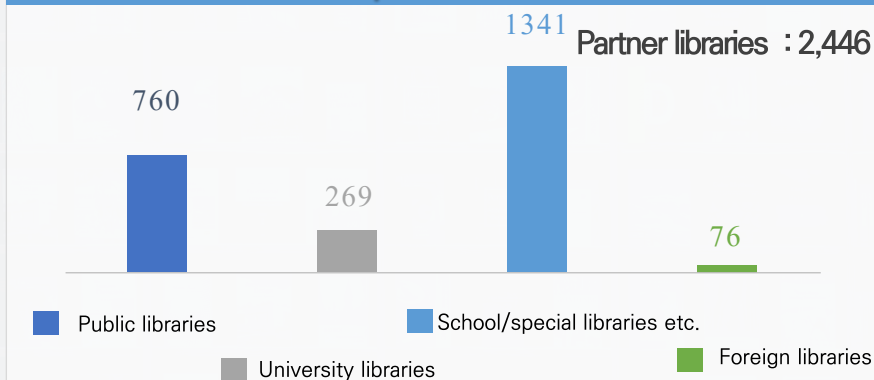


- Digitization project implemented since 1995
- Digitization of the NLK's old collection, general books, old newspaper, serials, culture/art materials
- 1,539,024 books digitized by the end of 2020(35% completed)

## 1,539,024 books digitization



## National knowledge resource sharing service used in partner libraries





## Advanced digital shooting equipment introduced

### Robot scanner(2)



- Resolution : 300-400dpi
- Scanning speed : 2,500 pages/hour
- Maximum size : 320 X 320mm
- Feature : high efficiency

### Scan studio



- Resolution : 100 million pixels, 9000dpi
- Maximum size : A1(841 X 594mm)
- Feature : high resolution, various form

# National collection digitization process

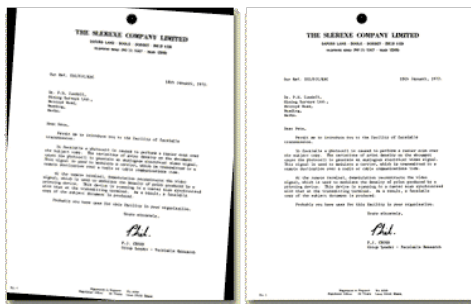
## 01 Image scanning



## 02 Table of Contents



## 03 Quality check



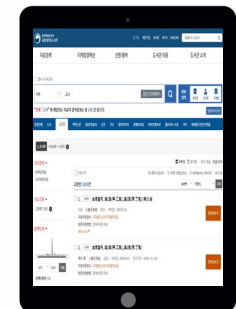
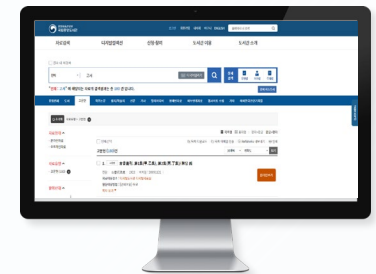
Before

After

## 04 OCR



Use the  
material



# National collection digitization process

## 01 Image scanning



- TIFF or JPEG
- preservation 400dpi
- service 300dpi
- Use purpose, scanning or photographing considering material condition



Class	Material	Speed	Use purpose
Scanning	<ul style="list-style-type: none"> <li>• Books, maps, single sheet document</li> </ul>	<ul style="list-style-type: none"> <li>• Fast(*damage concerns)</li> </ul>	<ul style="list-style-type: none"> <li>• Large scale digitization and service</li> </ul>
Photographing	<ul style="list-style-type: none"> <li>• Rare collection, oxidized books, picture</li> </ul>	<ul style="list-style-type: none"> <li>• Slow</li> </ul>	<ul style="list-style-type: none"> <li>• High resolution materials for exhibitions, research etc.</li> </ul>



## National collection digitization process

## 02

## Create Table of Contents



- Enter TOC information on corrected material
- Apply TOC input rule
- Describe the full-text as it is
- TOC check and link



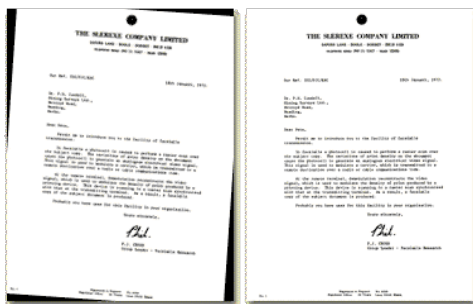
Order	Field name	Tag	Input example
1	Control number	Id	<id>KMO000000142
2	Call number	Call	<call>551-12-1(1973)
3	Total no. of pages	Count	<count>160
4	Year	Date	<date>
5	Publisher	Pub	<pub>
6	Book title	Title	<title>
7	TOC body	body	<body>TOC</body>

※ TOC tag structure

# National collection digitization process

## 03

## Quality Check



Before

After

- Check considering the feature and condition of scanned material  
※ old books, modern publications, newspapers etc.
- TOC creation guideline to provide quality information
- Full-text(original file) system linkage

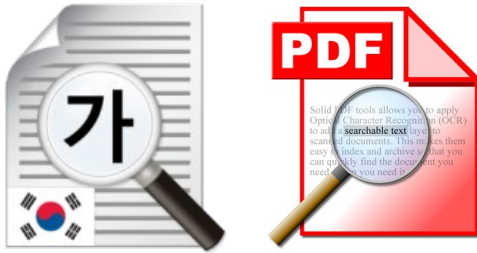


Class	Image	Page	TOC	Linkage
Content	<ul style="list-style-type: none"> <li>• Slope</li> <li>• Space</li> <li>• Noise</li> <li>• Image cut off</li> </ul>	<ul style="list-style-type: none"> <li>• Page omission</li> <li>• Image overlap</li> <li>• Page order</li> </ul>	<ul style="list-style-type: none"> <li>• TOC typo</li> <li>• TOC link</li> </ul>	<ul style="list-style-type: none"> <li>• System, upload and full-text linkage(original file)</li> </ul>

# National collection digitization process

## 04

## Character recognition(OCR)



- Full-text(original file) created : image file  
(text retrieval is not possible)
- Read character, symbol etc. on the scanned image  
※ image format → text format
- Language, paper quality and print condition affects recognition rate



Class	1. Full-text (original file) creation	2. OCR processing	3. PDF creation
Content	Image file, TOC	Image → text conversion	text PDF creation
Utilization	1. Enhance convenience with text retrieval 2. Research on specific material, big data analysis ※ Chinese character, vertical writing, old material recognition rate issue		

※ Research for OCR recognition rate improvement (2020) / Professor Wayne de Fremery



## Key success factors of national collection digitization



# Key success factors of national collection digitization

## Expand discovery of shared works

- Copyright-expired works
- Donated works
- Expand internet service by discovering orphan works etc.

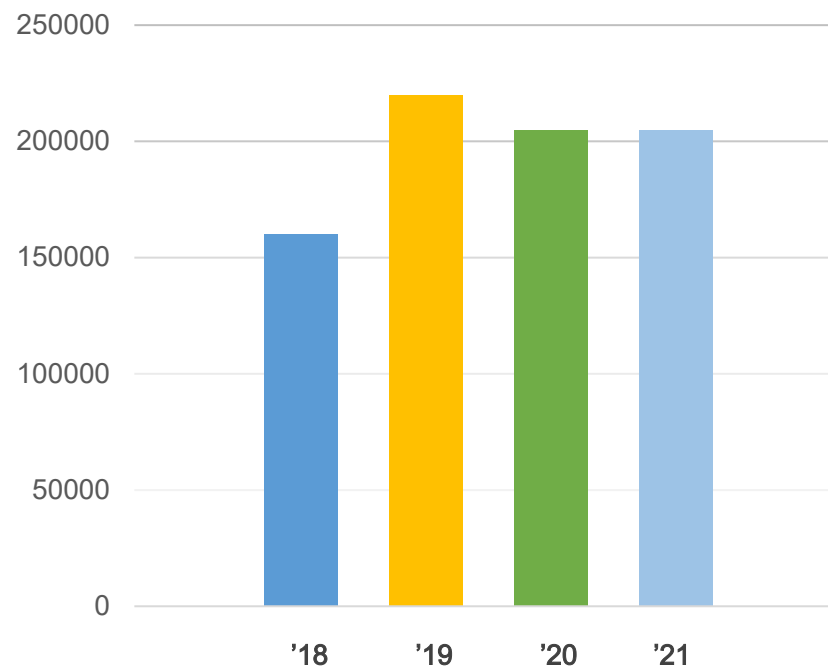
※ Copyright Act **Article 35 4**

(Replication by culture facilities etc.)

## Create database of texts that allow text retrieval

- Original image DB → text conversion
- Language information analysis through text mining

## About 4.27 million books(priority materials)



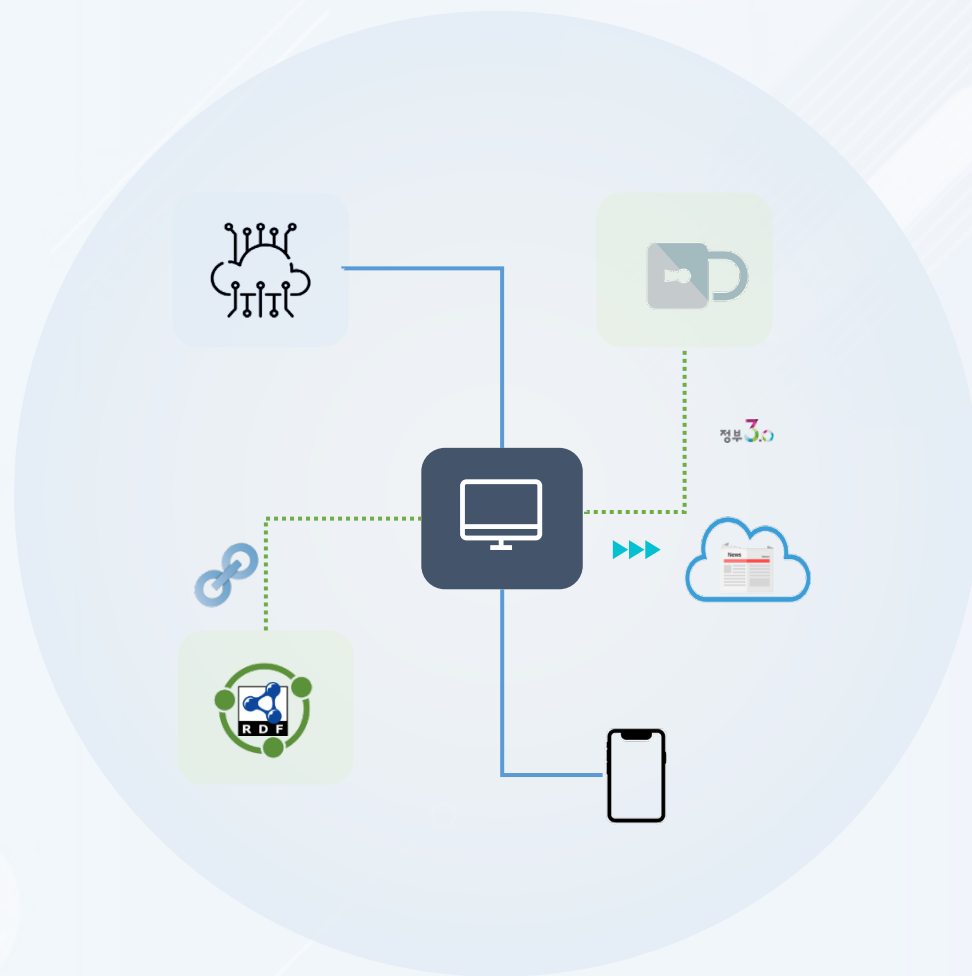
[Annual digitization status (books)]

# Key success factors of national collection digitization

## Database linkage/convergence

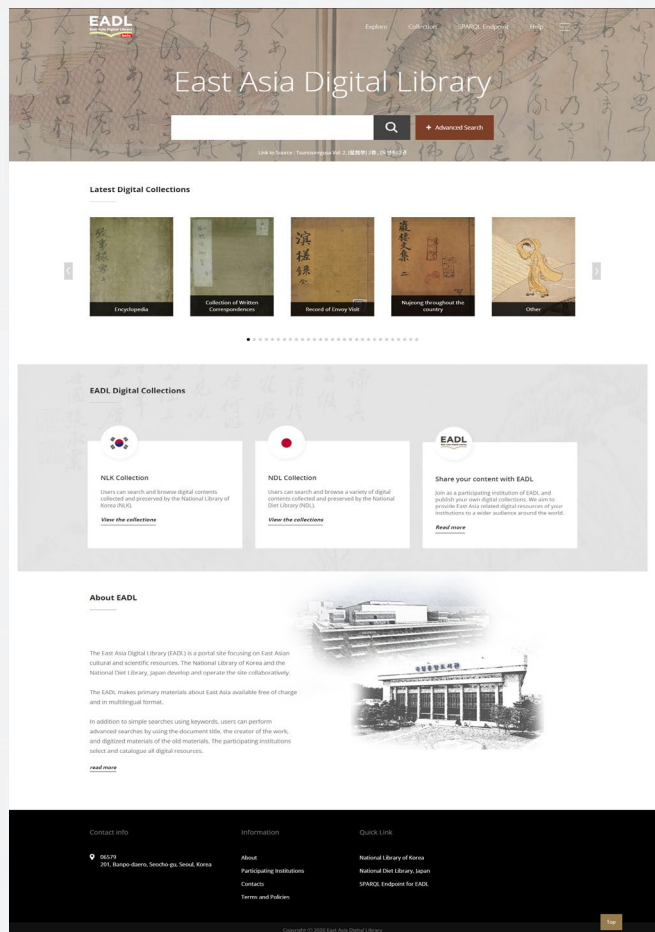
- Conveniently use and provide copyright-expired materials
- Fragmentary · segmental DB  
→ **integrate DB, link with relevant DB**  
(Korean Memory, Digital Jiphyeonjeon etc.)
- Link data by issuing LOD etc.  
(public-private, public-public data linkage etc.)

※ Enhancing metadata quality for data linkage is key





# Opening, sharing, expanding digitized materials



## Case. EADL(East Asia Digital Library, previously CJKDL)

- ✓ Integrated search and convenient access to East Asia's digital academic/culture resources
- ✓ Participating institutions  
National Library of Korea, National Diet Library of Japan
- ✓ Materials provided
  - Digital full- texts held by the NLK (119 titles), bibliographic data (3,926 items)
  - Digital full - texts held by the NDL (80titles), bibliographic data (3,988 items)
- ✓ Main service  
Integrated search, digital collection provision

# Opening, sharing, expanding digitized materials

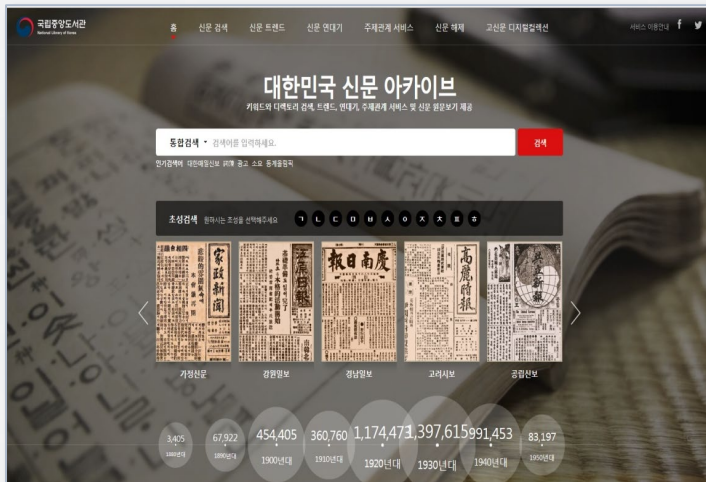
## Case. Korean Newspaper Archive

Rare and old newspaper DB-based

Integrated newspaper information system

✓ Anyone can use Korean rare and old newspaper anytime, anywhere

✓ Search by keyword, publication date/use



(<http://nl.go.kr/newspaper>)

# Opening, sharing, expanding digitized materials

## Case . Japanese Government General of Korea Gazette Utilization System

### Japanese Government General of Korea Gazette DB advancement and utilization system

- ✓ 11,235 volumes / 335,527 articles published by the Japanese Government General of Korea during the 36 years of Japanese rule(1910~1945)  
Search by date, classification, person's name

The screenshot shows the '조선총독부관보(1910~1945)' (Joseon Chongdokbu Gwanbo) search interface. The top navigation bar includes '디지털컬렉션' (Digital Collection) and '조선총독부관보(1910~1945)'. The left sidebar lists various collection categories like '주제별컬렉션' (Topic-based Collection) and '대한민국 신문 아카이브' (Republic of Korea Newspaper Archive). The main search area has tabs for '관보' (Gwanbo) and '오늘의관보' (Today's Gwanbo). Below these are search filters for '카테고리 검색' (Category Search), '일자 검색' (Date Search), '분류별 검색' (Classification Search), and '인명별 검색' (Name Search). The search results section shows a table with columns for '제목' (Title), '저자' (Author), and '일자' (Date). The table lists several documents, including '조선총독부관보(1910~1945)', '조선총독부관보(1910~1945)', '조선총독부관보(1910~1945)', '조선총독부관보(1910~1945)', and '조선총독부관보(1910~1945)'. The bottom of the page shows the URL 'https://nl.go.kr/NL/contents/N20302010000.do'.

(<https://nl.go.kr/NL/contents/N20302010000.do>)

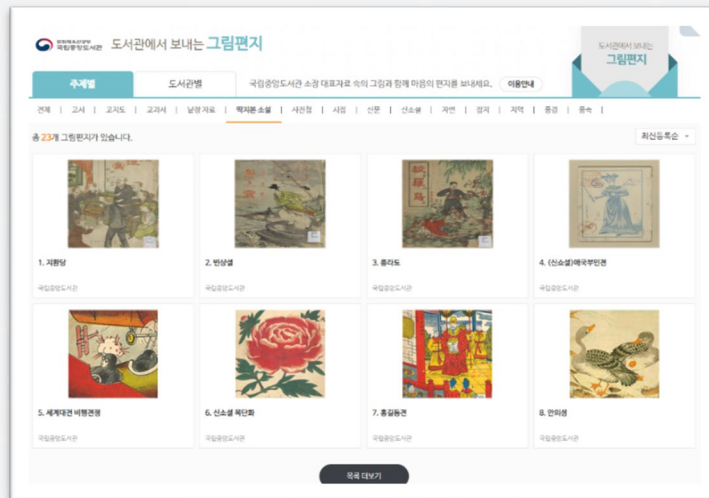


# Opening, Sharing and Expanding Digitized Materials

## Case . E-picture Card Service

### Utilization of the NLK's Collection

## E- picture card mail service



(<https://www.nl.go.kr/ecard>)

- ✓ Use covers and illustrations of the library's collection (without additional verification procedure)

Main function : download e- picture card, write and send letter

- ✓ Mobile service added in 2018

The background is a solid blue color with various abstract white geometric elements. On the left, there is a small triangle above a grid of dots. In the upper right, there are several parallel diagonal lines. On the right side, there are overlapping rounded rectangular shapes and a series of closely spaced parallel lines. In the lower left, there are more rounded shapes and diagonal lines. In the lower right, there is a single diagonal line ending in a small circle.

# Questions & Answers





Thank You